Jurnal Mandiri IT
Vol. 14 No. 4, April (2026), pp. 291-296

ISSN 2301-8984 (Print), 2809-1884 (Online)
Published by Institute of Computer Science

Handwritten text segmentation using deep learning method

Zulkarnaen Hatala[1], Ahmad Thariq[2], Josseano Parera[3], Muhammad Hatala[4]
[1,2,3]Informatics Department, Politeknik Negeri Ambon, Indonesia
[4]Industrial Engineering Department, Institut Teknologi Kalimantan, Indonesia

ARTICLE INFO

Article history:

Received Jan 19, 2026
Revised Feb 4, 2026
Accepted Apr 2, 2026

Keywords:

Deep Learning;
Document Image Analysis;
Handwritten Recognition;
Text Segmentation;
U-Net.
.

ABSTRACT

The rapid development of artificial intelligence and deep learning technologies has increased the risk of digital task fabrication in academic environments, encouraging educators to reintroduce handwritten assignments as an authentic evaluation method. In handwritten document analysis systems, background segmentation is a critical preprocessing step that separates text from complex document backgrounds. This study proposes the use of the U-Net deep learning architecture for background segmentation of handwritten document images. Two datasets were employed: the public cBAD dataset and a custom dataset consisting of Indonesian handwritten student assignments. Both datasets were processed using an identical pipeline and evaluated using 5-fold cross-validation. Model performance was measured using the Dice Similarity Coefficient and Intersection over Union (IoU). Experimental results show that the proposed U-Net model achieved an average Dice coefficient (F1-Score) of 0.74 on the cBAD dataset and 0.83 on the student assignment dataset. These results indicate that the model performs consistently and demonstrates stable generalization across cross-validation folds. Therefore, the proposed approach is suitable as an initial segmentation stage in handwritten document recognition systems.

Corresponding Author:

Zulkarnaen Hatala,
Informatics Department,
Politeknik Negeri Ambon,
M. Putuhena Street, Rumahtiga, Ambon, Maluku, 97234, Indonesia
Email: dzulqarnaenhatala@gmail.com

1 INTRODUCTION

The rapid advancement of artificial intelligence (AI) and deep learning technologies has significantly influenced academic assessment methods (Jo et al., 2020). The widespread availability of generative AI systems has increased concerns regarding digital assignment fabrication, prompting educators to adopt alternative assessment strategies such as handwritten assignments that possess unique biometric characteristics. Educators are encouraged to reintroduce handwritten assignments as an authentic evaluation method (Rabaev & Litvak, 2025).

Handwritten documents exhibit high variability due to differences in writing styles, stroke thickness, text orientation, and background textures (Mechi et al., 2019). These challenges are further amplified in scanned documents that suffer from noise, ink degradation, and paper artifacts. Conventional segmentation methods based on global or adaptive thresholding, such as Otsu (Otsu, 1979) and Sauvola (Sauvola & Pietikäinen, 2000), often fail to handle such variability consistently. Recent advances in deep learning, particularly convolutional neural networks (CNNs), have demonstrated superior performance in image segmentation tasks by learning complex feature representations directly from data (Calvo-Zaragoza & Gallego, 2019; Oliveira et al., 2018). The U-Net architecture (Lukács et al., 2025; Neha et al., 2025; Ronneberger et al., 2015), which employs an encoder–decoder structure with skip connections, has proven effective in preserving spatial information during segmentation.

Recent advances in document image analysis indicate that deep neural networks have become the dominant approach for handwritten text segmentation. In particular, encoder–decoder architectures are effective at modeling complex spatial features and heterogeneous background patterns (Tensmeyer & Martinez, 2017). Several studies have further shown that deep learning–based segmentation frameworks significantly outperform traditional thresholding and morphology-based methods when applied to both historical and modern handwritten documents, especially under noisy scanning conditions (He & Schomaker, 2019; Xiong et al., 2021).

In handwritten document image analysis, background segmentation plays a fundamental role in separating textual content from document backgrounds. Accurate segmentation is essential to ensure reliable performance in subsequent processes, including optical character recognition, feature extraction, and document digitization (Likforman-Sulem et al., 2007). Errors in the segmentation stage may propagate and degrade the overall recognition accuracy (Fizaine et al., 2024).

Although U-Net has been widely applied to international handwritten document datasets such as cBAD and READ-BAD (Diem et al., 2017), studies focusing on Indonesian handwritten documents remain limited and often rely on conventional preprocessing techniques. Existing works are largely fragmented and task-specific, often relying on conventional preprocessing pipelines or script-specific recognition models. Most available studies emphasize indigenous or regional scripts, such as Balinese (Prasetyo & others, 2022), Javanese (Arifin, 2017), rather than addressing handwritten documents in the Indonesian national language (Bahasa Indonesia) used in modern academic and administrative contexts. Recent efforts toward Indonesian document understanding have primarily focused on dataset construction and document extraction frameworks for local languages, highlighting the scarcity of robust, deep learning–based segmentation and recognition studies for contemporary Indonesian handwritten documents (Farhansyah et al., 2024). This research aims to address this gap by applying U-Net-based segmentation to Indonesian handwritten document images using robust evaluation strategies. The adoption of quantitative experimental protocols with cross-validation is therefore essential to ensure robust performance evaluation and generalization of segmentation models across diverse handwriting styles and document layouts. The combination of Binary Cross-Entropy (BCE) and Dice Loss was adopted to address the foreground–background class imbalance inherent in handwritten document segmentation. In such images, background pixels typically dominate, which may bias standard cross-entropy optimization toward the majority class. BCE provides stable probabilistic pixel-wise supervision, while Dice Loss directly optimizes region overlap and is less sensitive to class imbalance (Milletari et al., 2016). Integrating both losses enables balanced learning between pixel-level accuracy and global segmentation consistency, a strategy commonly applied in semantic segmentation networks such as U-Net (Long et al., 2015; Ronneberger et al., 2015).

Based on the identified research gap, this study contributes both technically and contextually to handwritten document segmentation. Although deep learning models such as U-Net have shown strong performance in document segmentation (Oliveira et al., 2018; Ronneberger et al., 2015), most prior studies focus on historical datasets such as cBAD and READ-BAD (Diem et al., 2017; Grüning et al., 2018) or degraded archival documents (He & Schomaker, 2019). Research on contemporary Indonesian handwritten documents remains limited, as existing Indonesian works mainly address regional script recognition or document extraction (Arifin, 2017; Farhansyah et al., 2024). Therefore, this study contributes by introducing and annotating an Indonesian handwritten student essay dataset for pixel-level segmentation. In addition, this research provides a contribution in purpose by positioning segmentation as a preprocessing stage for academic authenticity verification in response to AI-assisted task fabrication concerns.

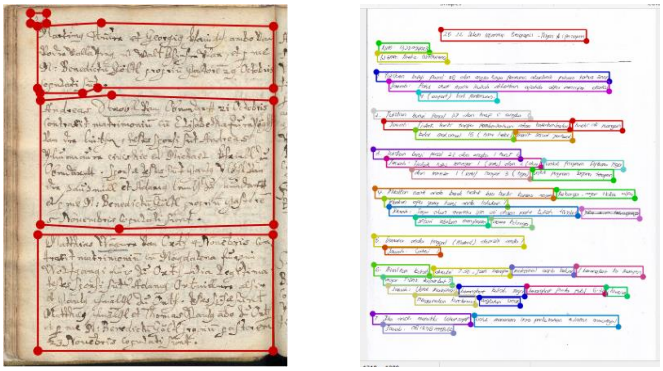
2 RESEARCH METHOD

This study employs a quantitative experimental approach using deep learning for handwritten document image segmentation.

Dataset

Two datasets were used: the publicly available cBAD dataset from ICDAR 2017 (Diem et al., 2017; Grüning et al., 2018) and a custom dataset consisting of scanned handwritten student essays written in Indonesian. Student essays are annotated in this research. Both datasets include

pixel-level ground truth masks distinguishing foreground (handwriting) and background. All handwritten student documents were used with explicit permission from the authors. The students provided informed consent allowing their handwritten data to be utilized and published for research purposes. Examples of both datasets are shown in Figure 1. Meanwhile, the paper size, scanning resolution, and language are presented in the Table 1.



(a) (b)

Figure 1. Page examples (a) cBAD (b) student essay

Table 1. Dataset properties

No.	Dataset	Dpi	Language	Page size
1	cBAD	300 dpi	English	A4
2	Student essay	300 dpi	Indonesian	A4

Preprocessing

All images were converted to grayscale and normalized to the range [0,1] to improve numerical stability during training (Goodfellow et al., 2016). The images were resized to a width of 1024 pixels while maintaining aspect ratio and divided into fixed-size patches of 256 × 1024 pixels. The patch size of 256 × 1024 pixels was selected to preserve horizontal text-line continuity while remaining computationally feasible. Handwritten documents are dominated by long horizontal lines; therefore, fixing the width at 1024 pixels ensures that each patch maintains complete line structures rather than fragmented segments. Preserving spatial continuity is important for segmentation networks such as U-Net, which rely on contextual feature learning (Long et al., 2015; Ronneberger et al., 2015). Patch-wise training is commonly used in document segmentation to balance context and hardware constraints (Mechi et al., 2019; Oliveira et al., 2018). Increasing patch height significantly raises GPU memory usage, so limiting it to 256 pixels ensures stable and efficient training.

Model Architecture

The segmentation model was built using the U-Net architecture, consisting of an encoder for feature extraction and a decoder for segmentation map reconstruction. Skip connections were applied to preserve spatial features (Ronneberger et al., 2015).

Training and Validation

The model was trained using the Adam optimizer due to its adaptive learning rate capabilities (Kingma & Ba, 2014). A combined Binary Cross-Entropy and Dice loss function was used to address class imbalance (Milletari et al., 2016). Model evaluation was conducted using 5-fold cross-validation (Kohavi, 1995).

Evaluation Metrics

Model performance was assessed using Dice Similarity Coefficient and Intersection over Union (IoU), which are more representative for segmentation tasks than accuracy alone (Long et al., 2015). The segmentation results were analyzed quantitatively using evaluation metric values and qualitatively through visual inspection of the correspondence between the segmentation

outputs and the reference masks. This analysis was conducted to assess the reliability of the U-Net model in separating handwritten text from the background and to identify remaining limitations.

3 RESULTS AND DISCUSSION

Results

Qualitative results can be observed in the segmentation of a representative sample from both databases. The segmentation results of the smallest image patches processed by the U-Net model are presented in Figure 2. These patches are then combined into original full-page documents as shown in Figure 3. Experimental results indicate that the proposed U-Net model achieved an average Dice coefficient of 0.74 on the cBAD dataset and 0.83 on the student essay dataset. The relatively low standard deviation on the Indonesian dataset demonstrates model stability and good generalization capability. The numerical values of F1-Score are presented in Table 2, while IoU is shown in Table 3. The quantitative evaluation also demonstrates that the proposed U-Net model achieves consistent segmentation performance across both datasets. On the cBAD dataset, the model obtained an average F1-score (Dice coefficient) of 0.7440, corresponding to an Intersection over Union (IoU) value of 0.5924. In contrast, the Student Essay dataset achieved a higher average F1-score of 0.8326, with a corresponding IoU of 0.7136. The improvement in both F1-score and IoU indicates superior overlap between predicted segmentation masks and ground truth annotations in the Student Essay dataset. Moreover, the lower standard deviation observed in the Student Essay dataset (0.0266) compared to cBAD (0.1666) suggests more stable and consistent model performance across cross-validation folds.

Table 1. F1-Score for both datasets

No.	Dataset	F1-Score	Std Dev
1	cBAD	0.7440	0.1666
2	Student Essay	0.8326	0.0266

(a) cBAD (b) student essay

Figure 2. Segmentation result of patches (a) cBAD (b) student essay

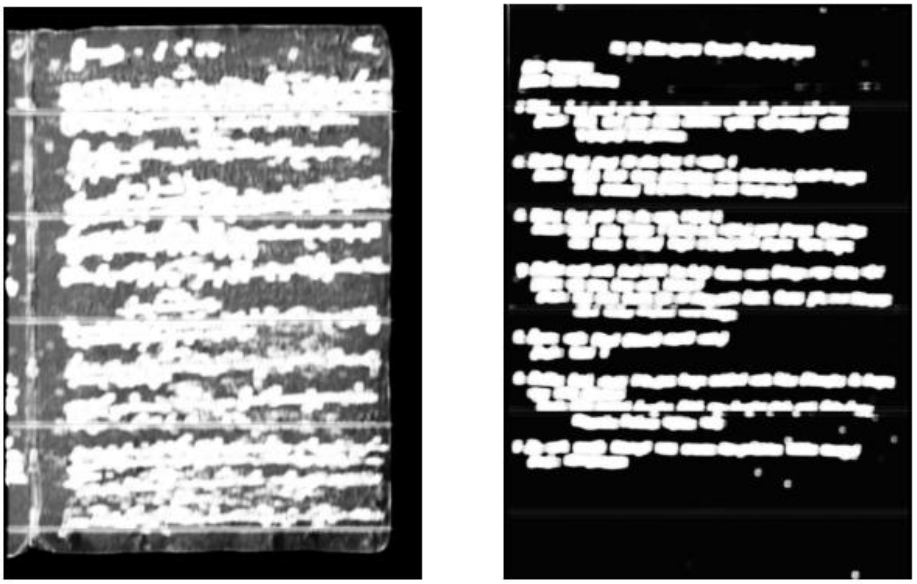
Table 2. IoU score of both datasets

No.	Dataset	IoU
1	cBAD	0.5924
2	Student Essay	0.7136

Discussion

Qualitative analysis reveals that some segmented regions contain merged text lines, indicating the need for additional post-processing stages such as line or word segmentation. In Figure 2 and Figure 3 The merged lines and words are more prevalent in the cBAD dataset, primarily because its ground truth annotations are less detailed compared to the student handwritten dataset. The observed discrepancy between the F1-score and IoU values is expected, as IoU is a more conservative metric that penalizes both false positives and false negatives more strictly. While the cBAD dataset achieved a moderate IoU of 0.5924, this value reflects the inherent complexity and variability of historical handwritten documents, as well as less detailed ground truth annotations. Conversely, the higher IoU of 0.7136 obtained on the Student Essay dataset indicates a more accurate separation between handwriting and background, which can be attributed to cleaner document conditions and more precise annotation quality. These findings confirm that although Dice coefficients provide an optimistic measure of segmentation overlap, IoU offers a stricter and more reliable indicator of segmentation robustness. Overall, the combined use of F1-score and IoU provides a comprehensive evaluation of the proposed segmentation model,

demonstrating its suitability as a preprocessing stage for handwritten document recognition systems.



(a) cBAD (b) student essay

Figure 3. Full page segmentation results (a) cBAD (b) student essay

4 CONCLUSION

This study demonstrates that the U-Net architecture is effective for background segmentation of handwritten document images, including Indonesian handwritten documents. The achieved Dice coefficients and low variance confirm the robustness and stability of the proposed model. The most promising direction for further research is the integration of transformer-based or hybrid CNN–transformer architectures to enhance global context modeling in complex document segmentation. While U-Net effectively captures local spatial features, transformer mechanisms can better model long-range dependencies and heterogeneous background patterns, which are common in degraded or densely written documents. Such architectures may significantly improve segmentation accuracy on complex handwritten pages with overlapping lines, irregular layouts, or noise. These results imply that the proposed U-Net segmentation framework can serve as a reliable preprocessing module for Indonesian handwriting recognition systems. By improving text–background separation, it enhances input quality for subsequent line and character recognition in contemporary Bahasa Indonesia documents.

REFERENCES

Arifin, M. (2017). *Handwritten Javanese Character Recognition Using Discriminative Learning* [Bachelor's Thesis, Universitas Indonesia]. https://doi.org/10.1109/ICITISEE.2017.8285521

Calvo-Zaragoza, J., & Gallego, A.-J. (2019). A selectional auto-encoder approach for document image binarization. *Pattern Recognition*, *86*, 37–47. https://doi.org/10.1016/j.patcog.2018.08.011

Diem, M., Kleber, F., Fiel, S., Grüning, T., & Gatos, B. (2017). cBAD: ICDAR2017 competition on baseline detection. *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, *1*, 1355–1360.

Farhansyah, M. R., Johari, M. Z. F., Amiral, A., Purwarianti, A., Yuana, K. A., & Wijaya, D. T. (2024). DriveThru: A Document Extraction Platform and Benchmark Datasets for Indonesian Local Language Archives. *arXiv Preprint arXiv:2411.09318*. https://doi.org/10.48550/arXiv.2411.09318

Fizaine, F. C., Bard, P., Paindavoine, M., Robin, C., Bouyé, E., Lefèvre, R., & Vinter, A. (2024). Historical text line segmentation using deep learning algorithms: Mask-rcnn against u-net networks. *Journal of Imaging*, *10*(3), 65. https://doi.org/10.3390/jimaging10030065

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. https://www.deeplearningbook.org

Grüning, T., Labahn, R., Diem, M., Kleber, F., & Fiel, S. (2018). READ-BAD: A Dataset and Evaluation Scheme for Baseline Detection in Archival Documents. *International Conference on Document Analysis and Recognition (ICDAR)*. https://doi.org/10.1109/ICDAR.2017.307

He, S., & Schomaker, L. (2019). DeepOtsu: Document enhancement and binarization using iterative deep learning. *Pattern Recognition*, *91*, 379–390. https://doi.org/10.1016/j.patcog.2019.01.025

Jo, J., Koo, H. I., Soh, J. W., & Cho, N. I. (2020). Handwritten text segmentation via end-to-end learning of convolutional neural networks. *Multimedia Tools and Applications*, *79*(43), 32137–32150. https://doi.org/10.1007/s11042-020-09624-9

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv Preprint arXiv:1412.6980*.

Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Proceedings of the International Joint Conference on Artificial Intelligence*, 1137–1145.

Likforman-Sulem, L., Zahour, A., & Taconet, B. (2007). Text Line Segmentation of Historical Documents: A Survey. *International Journal on Document Analysis and Recognition*, *9*(2–4), 123–138. https://doi.org/10.1007/s10032-007-0045-8

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully Convolutional Networks for Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965

Lukács, H. I., Beregi, B. Z., Porteleki, B., Fischl, T., & Botzheim, J. (2025). Attention U-Net-based semantic segmentation for welding line detection. *Scientific Reports*, *15*(1), 15276. https://doi.org/10.1038/s41598-025-00257-2

Mechi, O., Mehri, M., Ingold, R., & Amara, N. E. B. (2019). Text line segmentation in historical document images using an adaptive u-net architecture. *2019 International Conference on Document Analysis and Recognition (ICDAR)*, 369–374. https://doi.org/10.1109/ICDAR.2019.00066

Milletari, F., Navab, N., & Ahmadi, S.-A. (2016). V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. *Proceedings of the International Conference on 3D Vision (3DV)*, 565–571. https://doi.org/10.1109/3DV.2016.79

Neha, F., Bhati, D., Shukla, D. K., Dalvi, S. M., Mantzou, N., & Shubbar, S. (2025). An analytics-driven review of U-Net for medical image segmentation. *Healthcare Analytics*, 100416. https://doi.org/10.1016/j.health.2025.100416

Oliveira, S. A., Seguin, B., & Kaplan, F. (2018). dhSegment: A generic deep-learning approach for document segmentation. *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 7–12. https://doi.org/10.1109/ICFHR-2018.2018.00011

Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 62–66. https://doi.org/10.1109/TSMC.1979.4310076

Prasetyo, A. & others. (2022). DeepLontar: A Dataset for Handwritten Balinese Character Detection and Recognition. *Proceedings of an International Conference on Document Analysis*. https://doi.org/10.1038/s41597-022-01867-5

Rabaev, I., & Litvak, M. (2025). Recent advances in text line segmentation and baseline detection in historical document images: A systematic review. *International Journal on Document Analysis and Recognition*. https://doi.org/10.1007/s10032-025-00526-w

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Sauvola, J., & Pietikäinen, M. (2000). Adaptive Document Image Binarization. *Pattern Recognition*, *33*(2), 225–236. https://doi.org/10.1016/S0031-3203(99)00055-2

Tensmeyer, C., & Martinez, T. (2017). Document image binarization with fully convolutional neural networks. *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, *1*, 99–104. https://doi.org/10.1109/ICDAR.2017.25

Xiong, W., Zhou, L., Yue, L., Li, L., & Wang, S. (2021). An enhanced binarization framework for degraded historical document images. *EURASIP Journal on Image and Video Processing*, *2021*(1), 13. https://doi.org/10.1186/s13640-021-00556-4